Planning and conducting summative usability testing

E-GUIDE FOR TEST ADMINISTRATORS AND HFUE MANAGERS

By Niall Redmond, Industrial Design, Human Factors Engineering

Empowering Medtech Innovation®

StarFish MEDICAL 20 1999-2019

A GUIDE FOR TEST ADMINISTRATORS AND HFUE MANAGERS

There are a lot of things to consider when creating a summative usability test. The first time can be overwhelming. What scenarios do I test? How do I identify a critical task? How many participants do I need? Do I need to test the training?

In this article I will help you avoid common mistakes, like using a formative test structure for a summative usability test. I will also provide some tricks and tips to help you get the most out of your test participants and testing.

Contents

BEFORE YOU BEGIN TESTING

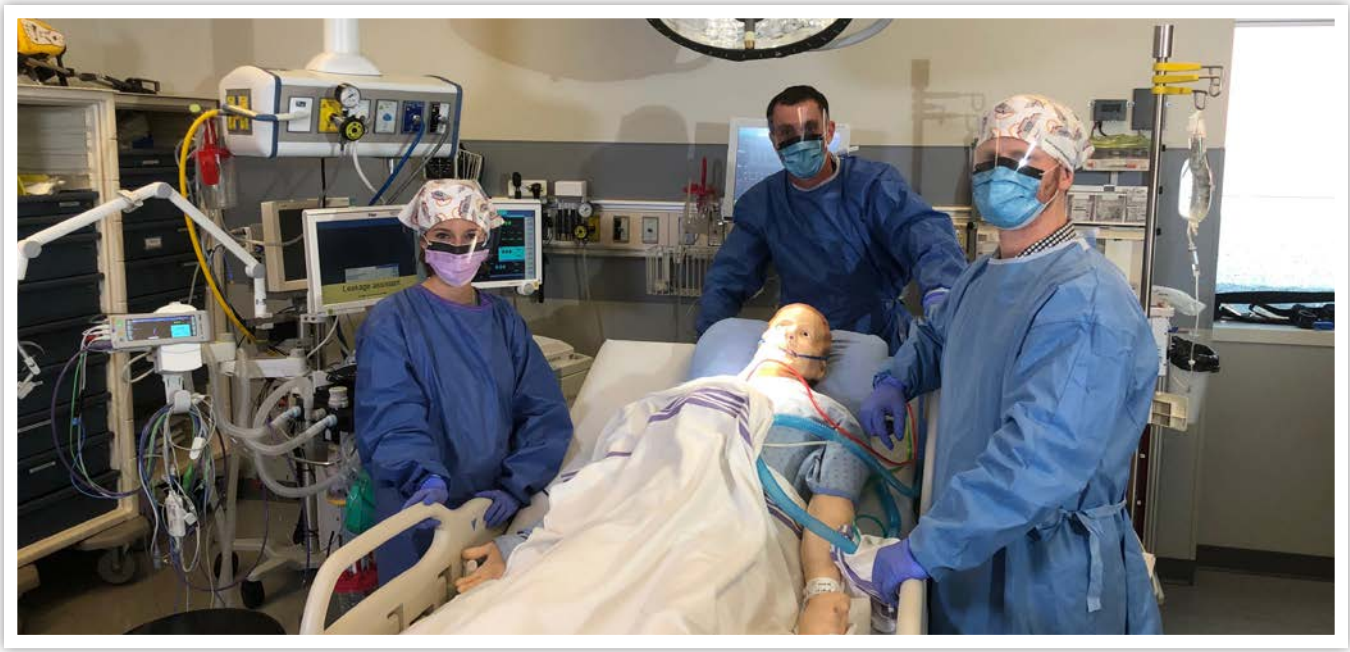
Institutional Review Board (IRB) approval

In general, the safest approach for research involving human subjects includes pre-approval by an Institutional Review Board (IRB) before research proceeds. Seeking prior IRB approval can be particularly important when the results are submitted the Food and Drug Administration (FDA). The FDA is part of the Department of Health and Human Services, which created the IRB process. IRB approval ensures that the research protects human subjects from undue risk.

An IRB approval is not always required. If the test is structured in such a way that participants are not exposed to physical hazards, it is acceptable to proceed without one. In this case, strict guidelines should be put in place to protect participants' identities and ensure that all aspects of testing an IRB review would have independently checked are identified and addressed in an appropriate manner. Aside from physical hazards, an IRB approval protects the rights and welfare of human research subjects by ensuring that physical, psychological, legal, and/or social risks to subjects are minimized, and when present, they justified by the importance of the research, and agreed to by subjects (informed consent). Nothing can be left to the researcher's judgment.

Protecting Human subjects

Be sure to identify and exercise all necessary IRB protections for human subjects. In some cases it might suffice to simply review these rights with the participant, including the right to withdraw from the test at any time without forfeiting compensation. In many cases it might also be necessary to keep a first aid kit on hand, maintain immediate access to a phone in order to call 911, or have a clinician available to treat a medical condition (e.g., engage an Emergency Medical Technician (EMT) or paramedic able to treat a severe asthma attack when testing an inhaler filled with a placebo).



TEST ENVIRONMENT

A summative usability test must be performed in a representative use environment with all environmental conditions accurately simulated. The ideal situation is to conduct testing in the actual use environment. Logistically this is not always feasible. If your use environment is a hospital, there are universities and facilities that rent out simulated operating rooms (OR) or care rooms when they are not in use. When setting up your own simulated environment, it is important to simulate all environmental conditions - from room size to props, equipment, furniture, temperature, lighting, noise and distractions - at the highest fidelity possible.

For example; using a hotel conference room as an OR is a possibility, but it is critical to book the correct hotel. If you are assessing a system with a cart, does the hotel floor have carpet? This would not be representative of an OR. Are there windows that will need to be boarded up? ORs generally don't have windows to the outside. Will there be hordes of tourists or school kids banging on the doors or creating unnatural distractions during the testing? How big should the room be? What should the lighting level be, and from what direction? Should there be a soundtrack with occasional alarms, beeps and whirring of equipment expected of an OR? What other equipment and furniture should be in the room?

The more accurate the environment, the more accurate the test results will be and participants will take the scenario seriously. I have found that when a test environment is set up loosely, participants are more self-conscious and acutely focused on the test. However, when the test environment is simulated with high fidelity, participants actively get into the role, even talking to the inanimate mannequin, asking how their day was, etc. This kind of active role play produces the most authentic and comparable results in usability testing.

No thinking out loud – This is not Formative Testing

Formative usability testing utilizes a perception, cognition, action (PCA) technique called "thinking out loud". PCA calls for the participant to verbalize his or her thoughts while performing a use scenario. Although not mandatory, it is suggested one employ this technique during formative usability tests. However, it is not appropriate to do so during a summative usability test. Regulatory bodies call for summative usability test participants to engage in a representative use environment. Requiring participants to think out loud would not be representative of a normal experience and could potentially change how participants interact with devices. That said, if a test participant tends to think out loud, then this is his or her normal behavior. The test administrator should neither encourage nor discourage the behavior. In other words, the participant can talk to himself or herself, but there should not be an associated dialog with the test administrator or testing staff.

Give an independent review period

If a user is likely to explore a device before using it, it is generally acceptable to designate a realistic and practical amount of time (say 15 minutes) at the beginning of a test session to reenact this process. For example, a person bringing home a new blood pressure cuff for the first time might read the user documentation before proceeding to use it.

That said, test personnel should not mandate that the participant take time to explore the device and its documentation, as not everybody would do this on their own volition. I, for instance, would open the packaging, put the instructions for use (IFU) to the side, stick blood pressure cuff on my arm, and start playing with it. If you were to mandate time with the device and user documentation, that would be considered additional training and not representative of actual use. Rather, it is appropriate to state something like the following:

"Imagine that you have just returned home from the pharmacy for the first time with your new blood pressure cuff. Review the blood pressure cuff and/or its instructions as you might at home when using a new device for the first time. Let me know when you feel comfortable using the device, and I will present some specific use scenarios."

TESTING PRESENTATION

Presenting "use scenario" and "knowledge task" prompts

"Use scenario" or "knowledge task" prompts are instructions provided to a participant that describe the high-level activity the participant should perform. For example, "Set up the device for use."

It is best to present prompts on a card for the participant to read aloud. Reading the prompt out loud has two functions, (1) it ensures the participant has read the entire prompt, and (2) it prevents the test moderator from consciously or subconsciously adding emphasis on parts of the task which they are interested in observing. This additional information could equate to assistance.

You may ask the participant to repeat the prompt in their own words. Repeating the prompt in their own words (not re-reading it) ensures that they understand the given use scenario or knowledge task. If there is any lack of understanding, the test administrator may choose to provide minor clarifications regarding the goal, taking great care not to convey information that equates to assistance.

Prompts should be worded in a manner that clarifies the activity's goal (e.g., starting after cleaning) but does not assist participants by guiding them on how to perform the tasks for a given use scenario (e.g., unscrew the cap). Avoid using terms that cue users to the proper menu selections or buttons to press. Furthermore, prompts should use words that all participants will understand, regardless of age and education level.

For example: *"Imagine you are at work and you need to set up a new patient to be ventilated. Perform all steps necessary to set up this new ventilator."*

Now compare it to this example of what not to do: *"Attach the four hoses to the ventilator, connect the hoses to the circuit's patient interface, program the ventilator to deliver a continuous rate of 25 mL/ Min, and then start the ventilation."*

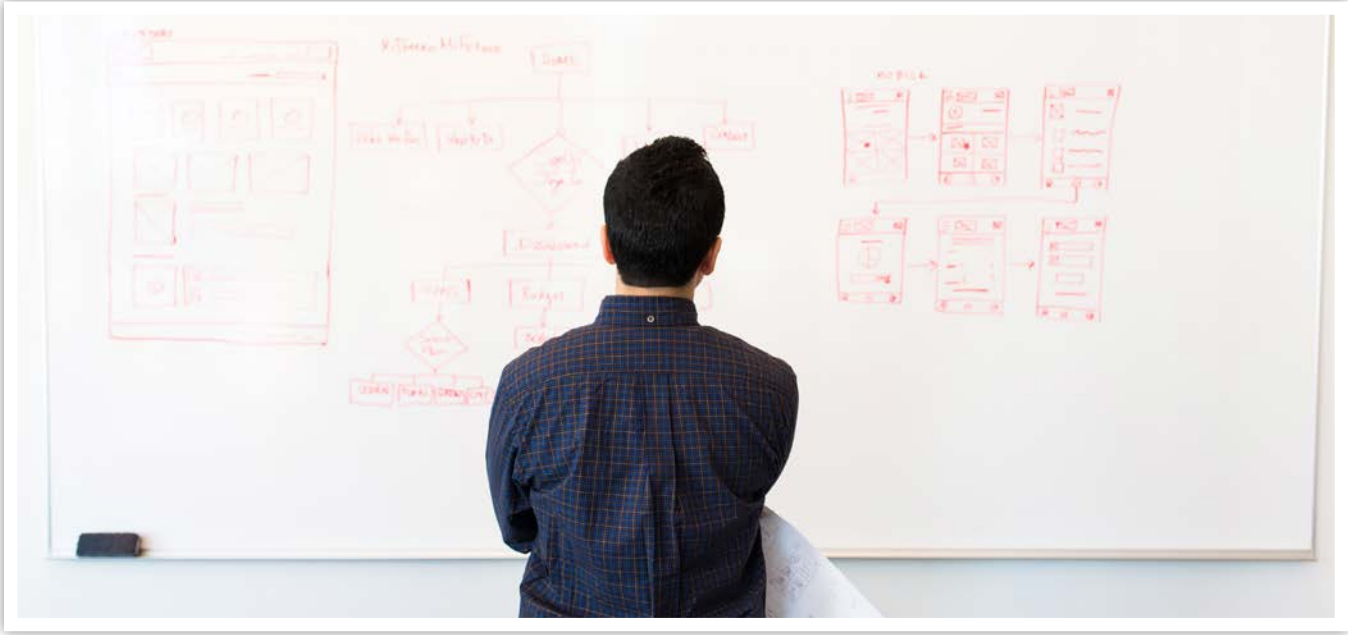
Presenting user documentation

It is important that a test administrator does not direct participants to follow the user documentation (e.g., instructions for use) because this would not occur in an actual use scenario. It is not appropriate to highlight the presence of user documentation. That said, it is acceptable to mention the user documentation in passing during the course of orienting the participant to a test environment. Perhaps mentioning it while pointing out equipment and other supplies in the test room.

Participants should use the user documentation only at their own initiative and only in cases where user documentation is available in the use environment. If user documentation is not normally available at the point of care, then it should not be present in the simulated testing. In situations where the user documentation would normally be available near the point of care (e.g., the user manual stored in a hospital office filing cabinet), test personnel may inform participants that user documentation is available upon request. The end goal is to present a realistic use scenario regarding accessibility and use of the accompanying documentation. It is required that user documentation content, as well as physical size, be production equivalent.

Natural workflow

It is important to keep the natural flow of a use scenario and not divide the use scenario into multiple steps or stop to pose questions or collect ratings. That said, it might be justifiable to divide a large span of activity that is normally separated into chunks by significant amounts of time (e.g., set up the device, clean the device, set up the device again after cleaning).



Timing of questions/interview

Do not pose questions to participants about use errors, close calls, and difficulties or their potential root causes until the final interview after all use scenarios and knowledge tasks have been performed. Asking such questions may constitute additional training and be unrepresentative of an actual use scenario. When test sessions are lengthy, it may be beneficial to ask a follow up question after each use scenario or knowledge task. An acceptable example

is: "How did that task go for you?" The question should be open-ended and nonspecific, but give the participant an opportunity to explain anything that might be forgotten by the time of the final interview. The participant response should be recorded by the moderator, but not addressed at that time. During the final interview, the test moderator should seek additional clarification on the participant's responses and root causes as needed.

TESTING TIPS

Test administrator script

Make sure to create a script for the test administrator to follow when conducting a test session. It may include an introductory blurb with initial, follow up, and final interview questions. You might also include an overview of each of the use scenarios and knowledge tasks, including the test session configuration, time critical functions and prompt for each. The script will include a lot of content found in the test plan, but presented in a condensed, special purpose format that facilitates administration of the test session.

Test administrator's assistance

Assistance from the test administrator is generally considered a test scenario fail. However, the test administrator may choose to provide assistance when the participant has reached an impasse and apparently (or by declaration) cannot continue independently. The test administrator may provide progressive levels of assistance, starting with something minimal, like referring the participant to the user manual. The test administrator may only assist a participant with a use scenario or knowledge task when the participant has attempted to complete the use scenario or knowledge task without assistance for a predetermined period of time (e.g., five minutes). That time should be sufficient for even slow performers to complete the task, if they will ever be likely to do so. Until that point in time, encourage the participants to persevere a while longer should they ask for assistance. All instances of a test administrator's assistance should be recorded and reported in the test report.

Subjective feedback

Recording participants' comments and feedback to questions asked by the test administrator during the final interview, provides an excellent set of subjective data that can be combined with objective data (e.g., observed use errors) to identify any patterns of concern.

Test personnel must exercise professional judgment when distinguishing between offhanded, limited subjective feedback and feedback or subjective observations that might constitute a pattern suggesting the need to report a use error, close call, or difficulty based on the identified pattern.

Identifying root causes

Ask questions during follow up and final interviews that generate responses which are helpful in determining the root causes of use problems which could lead to a use error, close call, or difficulty. Rather than allowing participants to blame themselves or answer "I don't know, I just did", encourage them to think about possible design-related root causes. "Was there anything about the device that may have led to this?" At a later point you can decide whether the participants' comments are accurate or inaccurate. This is perfectly acceptable.

Design updates

Do not look for design suggestions from participants during a summative test. Instead, summative usability testing should focus on collecting data that will determine if the device is usable, safe, and effective. The only appropriate time to look for design suggestions is at the end of the test. Then note if the participant says the design requires changes to ensure safe use.

When a helpline access is part of the system

If a helpline is available for a device during an actual use scenario, provide equivalent helpline access (actual or simulated) during the summative usability testing. The FDA qualifies a helpline as a risk mitigation. As such, validation is required. Provide a phone (equipped with a speaker) in the test room that participants can use to call an actual helpline or connect them to someone who will simulate the helpline service. Make sure that helpline calls are placed on speakerphone to allow test personnel to hear the dialog and audio/video recording equipment to capture the audio. Reflecting an actual use scenario, the user should have to figure out how to interact with the device under evaluation while handling the phone.

Do you think the device is safe?

Once the participant has attempted all use scenarios and knowledge tasks, it is suggested the test administrator give the participant an opportunity to assess the medical device's use safety. Regulatory guidance does not explicitly ask summative test participants to judge the overall use safety of a medical device. However, there is an expectation that the test administrator's final interview will focus on the device safety and ultimately expose whatever safety concerns might exist.

For example, the test administrator might ask: "Do you consider the ventilator safe to use 'as is,' or do you think it needs to be changed to ensure safe use?" In response to this particularly direct approach to assessing use safety, the participant should indicate whether they believe there is a need for design modifications to ensure the device is safe to use. Alternatively, concerns about use safety might arise in response to more general questions, such as: "What are your overall impressions of the device?" You may also seek participants' overall impressions of the clarity and effectiveness of the user documentation (e.g., user manual), if applicable.

TRAINING CONSIDERATIONS

Trained user groups

User groups should be identified based on their occupation, skills or specific device/user characteristics. This approach is usually correct. However, it might be appropriate to subdivide a particular occupation or skill set into two groups - one trained and one untrained. This particularly useful when you expect meaningful differences in how trained and untrained individuals approach use scenarios. For example, when evaluating the ability of a respiratory therapist to operate a ventilator, a manufacturer would expect the therapist to receive training before operating the ventilator. In reality, there may be situations that would call for an untrained respiratory therapist to operate a ventilator in basic ways.

Participant training

Participants should receive training if controls or risk mitigations are in place to ensure that all device users receive training. If the intended use of the device stipulates use by trained individuals, but such training is not assured, then the test should include a sample of both untrained and trained participants.

Representative training

Participant training should be delivered in a representative manner and expected time frame, using the content and instruction from the final training procedure. A manufacturer's employee can deliver the training even if the actual training might ultimately be delivered by a third party - such as a nurse who has been trained by the manufacturer. In all instances it is important that a representative level of separation be implemented into the test plan. You do not want end users getting a "gold standard" level of training from the manufacturer during testing, while real-world training will be provided by word of mouth and memory. For this reason, if the manufacturer's personnel are not delivering training during actual use, it is more realistic and simpler to take a "train the trainer" approach. For example, if you need to train people with diabetes to use a glucose meter, you could retain a consulting diabetes educator.

Training decay period

There should be a training decay period between the end of a training session and the actual test session. This shifts the test's focus from the participants' immediate recall of how to use the device to their ability to use the device based on retained knowledge from the training session along with any other relevant, pre-existing knowledge and skill. A decay period as short as one to two hours is sometimes acceptable, particularly if the training and test sessions are short and the expected gap between training and independent first use is short.

More commonly, longer training and test sessions are required, and substantial time might pass between training and independent first use. This gap suggests a longer training decay period is required (e.g., training on Day 1, testing on Day 2). When planning the decay period it is important to give each participant the same decay period. For example, the plan should state "training decay will be one day prior to testing." The plan should not say "training decay will be at least one day prior to testing." The FDA wants a standard amount of time for consistent results and to prevent some users receiving (online video) training one month before the testing while others receive the training 16 hours prior to the test. In the interest of simplified logistics, if the intended decay period is one day, it is generally acceptable for the actual decay to be more or less than 24 hours. For example, the training session may conclude on Day 1 at 5 p.m. and the test session may begin on Day 2 at 9 a.m., 16 hours later, rather than a full and exact 24 hours later.

Trainer observation

Individuals who provide training to participants during the summative usability test should not be involved with development of the use scenarios, knowledge tasks, or observe the summative usability test sessions. This ensures that the trainer will not be influenced by knowledge of the test method or results and consciously or subconsciously adjust their training approach in response.

TESTING AND RECRUITMENT

Test personnel during testing

During a test session it is acceptable for the test administrator and data analyst to either occupy the room with the participant or to direct the test from an adjacent control/observation room. A two way mirror may get more accurate results, but it is more common (and often desirable) for the test administrator and data analyst to occupy the room with the participant. This allows them to closely observe user interactions between the user and the device. It is important for the test moderator to maintain a good rapport with the participants. This will help the follow-up interviews. The data analyst should work in an unobtrusive manner. A possible exception is asking follow-up questions to augment or clarify those posed by the test administrator. Both the test moderator and the data analyst can make adjustments to the recording equipment to get a better view of the test being performed. (e.g., while the participant is cleaning specific components of the device, the moderator or analyst can adjust the view of a video camera to better capture the activity).

Recruitment

Create a screener script to guide recruitment of appropriate and qualified individuals for the usability test. Always over-recruit participants. Even if you do your own recruiting, approximately 10% to 20% of recruited participants will be disqualified or cancel their scheduled sessions (training and/or testing). This happens for any number of reasons, including no shows. For example, you should recruit 18 participants to ensure a 15 participant test. If you use a recruiter, make sure to consult with them well ahead of the usability test to determine the appropriate number to over-recruit for a given location and user population.

Competency test

People may lie, exaggerate or stretch the truth when excited about making money to test a product. To prevent the recruitment of "bad users" and corresponding bad test results, you may perform a check on a participant's ability to use a given medical device and accept or reject the individual based on competency test results. For instance, if an intended user needs to have previous training on an infusion pump as a requirement for the testing, an acceptable competency test would be to check that they know the basic functions and features of a given infusion pump. However, you should take this approach only when equivalent checks are expected in actual use (e.g., users receive mandatory training and/or must be certified to use the device).

Marketing questions

A summative usability test is a great opportunity to get your device in front of your target market and collect usability ratings or other marketing related data. Examples include ease of use, task and use performance speed, and other usability attributes of commercial interest. It is important not to cite the ratings as evidence of use safety, or include the data in the human factors and usability engineering report for FDA submission. If you do collect marketing related data, make sure to delay the collection of the data until after all other technical portions of the summative usability test are completed. This feedback should not be part of the main report. Instead, attach it as an appendix to the report.

FINAL THOUGHTS

Summative usability testing is no small task. The testing process is not only time consuming, but expensive. A lot of time is put into preparation, creating a summative usability test plan, conducting the testing, analyzing the data and writing the report. For this reason it is important to consider in full the information and recommendations above.

Successful summative testing requires that you plan user groups, recruit, select and train participants. Then you coordinate simulated use environments, IRBs and helplines if applicable.

Testing must be meticulously conducted to ensure the workflow is accurate, the wording of user scenario and knowledge task prompts are not leading, interview questions are open ended, looking at safety for root cause, and the environment and test administrators do not influence the test outcome. You must set up training in a representative manner and ensure the training decay is typical.

In summary, the additional effort put into planning carefully is invaluable. The result will be a test you can stand behind, something you are proud to submit to regularity bodies. It will also provide confidence that you are putting the safest product possible out into the marketplace.

Sources:

- *ANSI/AAMI HE75:2009/2013 Human factors engineering – Design of medical devices*
- *FDA, Applying Human Factors and Usability Engineering to Medical Devices, February 3, 2016*
- *AAMI Writing Human Factors Plans & Reports for Medical Technology Development, 2018*
- *IEC/TR 62366-2:2016, Medical devices -- Part 2: Guidance on the application of usability engineering to medical devices*
- *Image credits: Pexels.com and StarFish Medical*

Visit StarFishMedical.com for more Medical Device Human Factors, Industrial Design, and Usability blogs, videos, and insights.

PHONE: (250) 388-3537 | **TOLL FREE:** 1 (877) 822-3537

info@starfishmedical.com

StarFish MEDICAL 20 1999-2019